Cosmic Microwave Background Analysis Tools (COMBAT) II Final Report

Julian Borrill
Computational Research Division, Lawrence Berkeley National Laboratory
& Space Sciences Laboratory, UC Berkeley.

The Cosmic Microwave Background Analysis Tools (COMBAT) project's goals were to develop massively parallel implementations of CMB data analysis algorithms and to make them available to the CMB community. The final product here is the Microwave Anisotropy Dataset Computational Analysis Package (MADCAP) which includes tools for most of the key steps in the CMB data analysis chain:

- MADnes - derives the piecewise stationary noise correlation properties of a CMB time-ordered dataset and fills timestream gaps with an appropriately constrained noise realization in order to provide continuous data for subsequent fourier analysis.
- MADpre - generates the trivially invertable white-noise approximation to the full pixel-pixel noise correlation matrix which is used both to identify unresolvable pixels in CMB polarization data sets, and as a conjugate gradient preconditioner in map-making codes including MADmap.
- MADmap - uses by preconditioned conjugate gradient methods to calculate the maximum likelihood microwave sky temperature and polarization maps given the noise correlation properties determined by MADnes, the preconditioner from MADpre, and an arbitrarily complex pointing matrix.
- MADping - uses explicit dense symmetric matrix inversion to construct the full pixel-pixel noise correlation matrix (and, incidentally, the exact maximum likelihood sky maps) from the noise correlation properties determined by MADnes and an arbitrarily complex pointing matrix.
- MADspec - calculates the maximum likelihood CMB temperature and polarization auto- and cross-spectra from the maximum likelihood maps and their associated pixel-pixel noise correlation matrix generated by MADping.

To ensure the relevance of the development path, and to demonstrate the applicability of the MADCAP tools, the COMBAT team simultaneously applied them to real and simulated leading-edge CMB data sets:

- BOOMERanG NA, 98 & MAXIMA : The drivers for much of the early temperature-only MADCAP development, resulting in the first detections of a series of acoustic peaks in the CMB temperature anisotropy.
- BOOMERanG 03 & MAXIPOL : The successors to the above, these experiments drove the development of the later polarization capabilities of MADCAP. Their adoption of very different approaches (using pairs of polarization-sensitive bolometers and a rotating half-wave

plate respectively) ensured that the MADCAP tools retained breadth of applicability in this emerging field.

- Planck : The 3rd generation CMB satellite mission whose simulations already generate the largest CMB data sets to date, Planck has driven the development of MADCAP capabilities handling exceptionally large data sets across a range of parallel platforms, from small clusters to the largest supercomputers.
- EBEx & PolarBear : Examples of next-generation CMB polarization experiments that will use thousands of multiplexed detectors in place of the traditional tens of individual ones in order to achieve the signal-to-noise necessary to measure in detail B-mode polarization anisotropies, these experiments have driven in particular the development of on-the-fly reconstruction of a particular detector's pointing from the overall focal plane pointing, essential for the requisite compression of pointing data for such multiplexed sytems.

During the development of these tools it became apparent that the necessary increase in the size of CMB data sets meant that high performance computing (HPC) was going to become increasingly important to their data analysis. A spin-off of this work has therefore been the establishment of an allocation of HPC resources at the DOE's National Energy Research Scientific Computing (NERSC) Center that is shared by the CMB community, at any time supporting O(100) data analysts from O(10) experiments with major allocations of both supercomputer cycles and storage. Building on this, the CMB data analysis community's future HPC requirements were also highlighted in the most recent NERSC Greenbook. This shared resource has also proved to be an effective way of making the MADCAP tools available to people, as pre-installed packages on all of the NERSC machines.

In addition, the close relationship with NERSC has led to the development of MADbench – a benchmarking version of MADspec that is used to measaure the performance of supercomputers under the stresses of real scientific codes – which has been included in the vendor benchmark package for the procurement of the last 3 NERSC systems. MADbench has also been used to perform cross-platform comparisons, including overall integrated system and specific IO subsystem performance tests on a range of machines including NASA's Project Columbia, the Earth Simulator, all of the NERSC systems, two different Blue Gene systems, a Cray XT3, and a range of high-end clusters.

Finally, this success of this project in developing HPC for CMB data analysis directly fed into the recent tri-agency Weiss "Taskforce on Cosmic Microwave Background Research" , where high performance computational science research and development was identified a one of the 5 technical findings, and the leading role of the NERSC-based program in particular was called out.